

Assignment – Nanos.ai

Problem Statement:

Given the url of webpage like: <https://www.nanos.ai> and list of products, your task is to extract the most relevant words from the webpage to the given list of products/services.

Approach:

My approach to this given problem is by using word embeddings to find similar relevant keywords to the given list of product/services names.

1. Web Scraping:

- Scrape the webpage of the given url and saving it to a file.

```
<!-- Search Engine Optimization by Rank Math - https://s.rankmath.com/home -->
<title>The Online Marketing Hero For Every Business - Nanos.ai</title><link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Roboto%3A400%2C600%2C700%7CPoppins%3A400%2C400italic%2C500%2C600%2C700%7CRoboto%20Condensed%3A400%2C600%2C700%7CPoppins%3Aregular%2C700%7COpen%20Sans%3Aregular%2C700&#038;display=swap" /><link rel="stylesheet" href="https://nanos.ai/wp-content/cache/min/1/37e7a05f7722fc0a3decd4a9c8898c83.css" media="all" data-minify="1" />
<meta name="description" content="Nanos is the online marketing superhero for every business. From a large business to a small business the goal is the same: to grow!">
<meta name="robots" content="follow, index, max-snippet:-1, max-video-preview:-1, max-image-preview:large"/>
<link rel="canonical" href="https://nanos.ai/" />
<meta property="og:locale" content="en_US">
<meta property="og:type" content="website">
<meta property="og:title" content="The Online Marketing Hero For Every Business - Nanos.ai">
<meta property="og:description" content="Nanos is the online marketing superhero for every business. From a large business to a small business the goal is the same: to grow!">
```

2. Data cleaning/preprocessing:

- Cleaning data by removing web tags, punctuation marks, stopwords
- Converting words to lowercase
- Tokenizing words
- Converting umlauts characters to different format for german language

For ex: ä : ae; ö = oe; ü = ue

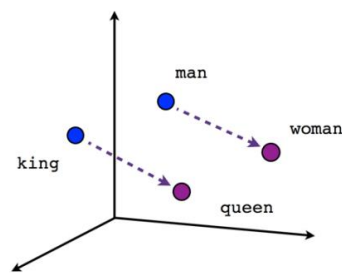
['online marketing hero every business nanosai skip content nanos digital advertising home us case studies ai engine blog videos sign nanos doityourself payasyougo advertising tool create optimized ad campaigns google facebook instagram one place nanos patented artificial intelligence technology start nowwatch works credit card required become marketing agency easy steps website facebook business page required online marketing simple like agency pocket browser cannot play provided video file step sign nanos tell us words great product service browser cannot play provided video file step nanos ai suggest ad text keywords interests platforms ad browser cannot play provided video file step decide budget want spend long run ads browser cannot play provided video file step log dashboard watch nanos ai action optimizing lightning speed seen start credit card required digital marketing superhero every kind business whether physical store online venture goal grow business understand existing marketing tools may complicated expensive everyone marketing expert afford hire one website landing page got covered build free landing page browser cannot play provided video file browser cannot play provided video file winwin solution nanos take confusion online advertising nanos fees transparent charge total advertising budget pause cancel ad campaign time remaining funds get refunded account automatically nanos proven track record bringing new clients business like restaurants auto sales real estate agencies video game developers online courses wedding planners start nowbook demo credit card required nanos agencies running small midsize agency looking tool optimize clients budgets cost

Assignment – Nanos.ai

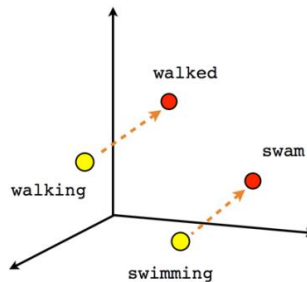
['gitarrenttotal zurich gitarren total zurich beware of the axe guitar service mitten zuerich bietet gitarren total professionelle ll zuverlaessigen reparaturservice fuer baesse akustische elektrische gitarren kleiner checkup grosse operation untersuchen l iebste body saiten holen klaenge instrument niemals darin vermutet haettest reparaturen service halsbruch winterschaden ganz no rmale abnutzungserscheinungen buende abrichten neubundierung pus wechseln intonation halsstab einstellen ausstattung unserer we rkstatt erfahrung erlauben einfachen service anspruchsvollen reparatur gerne kostenvoranschlag neubundierung beispielbarkeit ins trumentes haengt wesentlich qualitaet neubundierung ab verwandeln abgekaempfte lieblingsgitarre makellos beispielbares instrumen t besondere wuensche speziellen bunddraht kompensierten griffbrettradius finessen gehen gerne lackierung lack ab lackschaeden f alsche lackwahl lackiertechnik aesthetisches problem lack wesentlich einfluss klang unserer lackierwerkstatt bieten saemtlich e lackierverfahren hervorragender qualitaet alten schellack french polishlackierung ueber nitro dd polyesterlackierungen kleinere unfaele fachgerechten retouche unsichtbar custom shop traumst heissen kurven ganz normal traumst heissen kurven saiten fall fuer custom shop planen bauen individuelles instrument schrecken geheimsten wuenschen zurueck natuerlich fuer kleinere umbauten pu mechanik tremolowechsel restauration alte liebe paar risse bekommen lassen instrument neuem glanz erstrahlen erklingen fehlt ersatzteil fuer vintageinstrument beschaffen originalteil finden wege originalzustand moeglichst nahe kommen buy now contact us email this page created with huubsh create your free landing page today huubshcom']

3. Using pretrained word embeddings from spacy

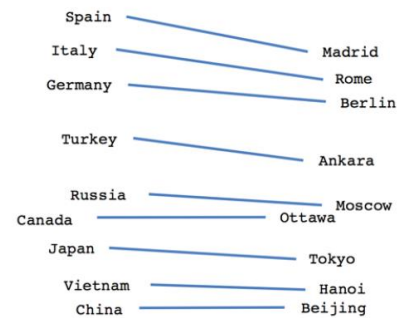
- Spacy has a package “en_core_web_lg” which is trained on Wikipedia corpus. and “de_core_news_lg” which is trained on news articles corpus.
- Word embedding is a learned representation for text where words that have the same meaning have a similar representation.



Male-Female



Verb tense



Country-Capital

- Words which are closer in distance (Euclidean or cos-sim) are similar and related.

Other Approaches:

There are always many other solutions to a given problem.

We can even use *Topic modeling*, *NER(Named Entity Recognition)* to train a model.

Since there is only data of a single webpage, these approaches won't work very well.